

6169-186
BOC9-2000-0052

**SPEECH PROCESSING BOARD
FOR HIGH VOLUME SPEECH PROCESSING APPLICATIONS**

Inventors: Harry W. Printz
Bruce A. Smith

International Business Machines Corporation

IBM Docket No. BOC9-2000-0052
IBM Disclosure No. BOC8-2000-0067

Express Mail Label No. EL740156380US

BACKGROUND OF THE INVENTION

Technical Field

The present invention relates to speech recognition and more particularly to a speech processing board.

State of the Art

Present modes of communication are rapidly changing due to the integration of the computer and telephone. Computer Telephony (CT) represents this integration and includes the utilization both of speech recognition technology and text-to-speech (TTS) technology. Companies such as International Business Machines Corporation have implemented telephone speech recognition platforms capable both of continuous speech recognition and TTS playback. As a result, CT has become one of the fastest growing applications markets for speech recognition, with many companies producing products specifically for the CT market.

For instance, Dialogic Corporation of Parsippany, New Jersey USA has developed a speech processing board for use in CT. Specifically, the Dialogic speech processing board is a CT solution which conforms to the compact PCI (cPCI) communications specification. The speech processing board can include an open architecture which can accommodate the integration of CT related resources such as a automatic speech recognition, TTS playback and call control. The architecture further can include a high-level application programming interface (API) based on the well-known Enterprise Computer Telephony Forum (ECTF) API. The Dialogic speech

processing board can include a CT Bus for facilitating the integration of the speech processing board with a CT system. The CT Bus in the Dialogic speech processing board is a time division multiplexing (TDM) bus that provides 1024, 2048, or 4096 time slots for exchanging voice, fax, or other network resources on the cPCI backplane.

5 Notably, the CT Bus conforms to the H.110 standard which allows CT application developers to build large, distributed, open CT systems in public network and customer premises environments.

By comparison, Lucent Technologies, Inc. of Murray Hill, New Jersey USA also manufactures a cPCI compliant speech processing board for use in CT applications. Lucent's speech processing board enables service providers to provide customers with speech-enabled applications in the CT environment. Like the Dialogic speech processing board, the Lucent speech processing board can support more than one hundred audio channels. Moreover, the speech processing board can support multiple speech applications such as speech recognition and TTS playback. Notably, the
10
Lucent speech processing board can provide flexible speech recognition capabilities ranging from simple connected digits to complex grammar-based, continuous speech. Finally, like the Dialogic speech processing board, the Lucent speech processing board meets the ECTF cPCI standards, including the industry-standard H.110 interface.
15

Still, as the volume of speech processing applications increases in a CT system,
20 both the Dialogic and Lucent speech processing boards are unable to adequately process each speech processing task using one speech processing board alone. In consequence, both Dialogic Corporation and Lucent Technologies, Inc. suggest the use

of multiple speech processing boards to handle high volume speech applications. The use of multiple speech processing boards, however, can consume valuable bus slots and can increase the number of hardware resources necessary to accommodate each speech processing board. Hence, what is needed is a speech processing board which

5 is optimized for high volume speech processing applications.

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
22

SUMMARY OF THE INVENTION

The speech processing board of the present invention is an optimized speech processing board for use in high volume speech processing applications. The speech processing board can include multiple processors, each which can execute multiple instances of speech applications for performing both large and small vocabulary recognition tasks. The speech processing board design and associated firmware can work in concert to provide state-of-the-art speech recognition capabilities for deployment in classical computer telephony (CT) applications or in gateways/endpoints of voice over IP (VoIP) applications. The speech processing board also can accommodate multiple instances of Text-to-Speech (TTS) applications. Finally, the speech processing board can support various levels of session control applications such as dialog manager natural language understanding (NLU) engines and traditional interactive voice response (IVR) applications.

A speech processing board configured in accordance with the inventive arrangements can include multiple processor modules, each processor module having an associated local memory, each processor module hosting at least one instance of a speech application task; a storage system for storing speech task data, the speech task data including language models and finite state grammars; a local communications bus communicatively linking each processor module through which each processor module can exchange speech task data with the storage system; and, a communications bridge to a host system, wherein the communications bridge can provide an interface to the

local communications bus through which data can be exchanged between the processor modules and the host system. Notably, the host system can be a CT media services system or a VoIP gateway/endpoint.

Each processor module can include a central processing unit (CPU) core having at least one memory cache which can be accessed by the CPU core; a processor bridge communicatively linking the CPU core to the local communications bus; and, a memory controller through which the CPU core can access the local memory, wherein the memory controller can be linked to the CPU core through a processor local bus. Additionally, a language model cache can be disposed in the local memory. Finally, a finite state grammar table can be disposed in the local memory.

The storage system can include a fixed storage device accessible by the processor modules through the communications bridge, wherein the fixed storage device stores active language models and finite state grammars used by the speech application tasks hosted by the processor modules; a commonly addressed language model cache, wherein the language model cache can store at least one image of a language model stored in the fixed storage device, each processor module accessing the language model cache through the communications bridge at a common address; and, a boot memory storing initialization code, wherein the boot memory is communicatively linked to the processor modules through the communications bridge, each processor module accessing the boot memory during an initial power-on sequence.

The local communications bus can be a PCI bus. More particularly, the PCI bus can be a 64-bit, 133MHz PCI bus. Alternatively, the PCI bus can be a 64-bit, 66MHz PCI bus. The communications bridge can include a PCI-to-PCI bridge having a PCI interface to the host system and an interface to an H.1x0 bus. The communications bridge also can include a processing element for managing message communications between the speech processing board and the host system according to a messaging protocol provided by the host system. Notably, the communications bridge can be implemented in a field programmable gate array (FPGA).

The speech processing board also can include a serial audio channel communicatively linking the processor modules to the communications bridge. The serial audio channel can provide a medium upon which audio data can be exchanged between individual processor modules and the communications bridge. An audio stream processor also can be provided which can be coupled to the communications bridge. The audio stream processor can be configured to extract audio information received in the communications bridge, store the extracted audio information and distribute the audio information over the serial audio channel to selected ones of the processor modules based on hosted instances of speech applications in each processor module.

In one particular embodiment of the present invention, a speech processing board can include multiple processor modules in the speech processing board; a PCI-to-PCI bridge interfacing the local PCI interface to a host CT system, a local PCI interface linking each processor module to the PCI-to-PCI bridge; a fixed storage

communicatively linked to the PCI-to-PCI bridge and accessible by the processor modules through a drive controller; a language model cache communicatively linked to the bridge; and, a boot memory communicatively linked to the bridge, the boot memory storing initialization code. Notably, the PCI-to-PCI bridge can include interfaces to an H.1x0 bus and a PCI bus.

A high-volume speech processing method in accordance with the inventive arrangements can include the steps of loading and executing a plurality of speech application tasks in selected ones of multiple processor modules in a speech processing board; loading in a commonly addressed storage separate from the multiple processor modules, selected language models for use by the speech application tasks; receiving audio data over an audio channel and distributing the audio data to particular ones of the processor modules, wherein the distribution of the audio data to particular ones of the processor modules is determined based upon speech application tasks executing in the particular ones of the processor modules; processing the received audio data in the particular ones of the processor modules using the language models selected for use by the speech application tasks; and, caching in the selected ones of the multiple processor modules portions of the selected language models used by the speech application tasks. The method also can include the steps of collecting speech task results from the selected ones of the multiple processor modules; and, forwarding the collected speech task results to a host CT system over a host communications bus.

BRIEF DESCRIPTION OF THE DRAWINGS

There are shown in the drawings embodiments which are presently preferred, it being understood, however, that the invention is not limited to the precise arrangements and instrumentalities shown, wherein:

5 Figure 1 is a block diagram illustrating a speech processing board configured in accordance with the inventive arrangements.

Figure 2 is a block diagram of a processing module for use in the speech processing board of Figure 1.

Figure 3 is a schematic illustration of the speech processing board of Figure 1 integrated with an ECTF-compliant computer telephony system.

DETAILED DESCRIPTION OF THE INVENTION

I. Overview

The present invention is a speech processing board which has been optimized for use in high volume speech processing applications. Unlike conventional speech processing boards, the speech processing board of the present invention can include multiple processor modules each of which can execute multiple instances of full function, large vocabulary speech recognition tasks similar to those of a conventional speech recognition engine with shared memory. The speech processing board can be deployed both in a conventional computer telephony (CT) architecture and in a voice over IP (VoIP) gateway/endpoint architecture. The speech processing board of the present invention also can accommodate multiple instances of text-to-speech (TTS) application tasks and small vocabulary speech recognition tasks.

Figure 1 is a block diagram illustrating a speech processing board 100 configured for use in high volume speech processing applications according to the inventive arrangements. The speech processing board can include multiple processor modules 102, a local communications bus 104, a storage system 106 and a communications bridge 108. Each processor module can have an associated local memory and can host therein one or more instances of selected speech application tasks. Speech application tasks can include both large and small vocabulary speech recognition tasks, speech synthesis (TTS) tasks, natural language processing and the

like. Each processor module 102 further can be communicatively linked with the local communications bus 104.

Processor modules 102 can exchange speech task data with the storage system 106 through the local communications bus 104. In one aspect of the present invention, the storage system 106 can include fixed storage 106A, a language model cache 106B and boot memory 106C. The fixed storage 106A can be a compact fixed disk drive analogous to a hard disk drive. The Microdrive® manufactured by International Business Machines Corporation of Armonk, New York USA is an example of compact fixed storage. The fixed storage 106A can store active language models and finite state grammars used by the speech application tasks in the processor modules 102. The processor modules 102 can access the fixed storage 106A through a disk controller such as an IDE or ATA compatible interface which is linked to the communications bridge 108.

By comparison, the language model cache 106B can be volatile or non-volatile memory such and can store at least one image of a language model stored in the fixed storage. As in the case of the fixed storage 106A, each processor module 102 can access the language model cache 106B through the communications bridge 108. Notably, the language model cache 106B can be accessed by each processor module 102 at a common address. Finally, the boot memory 106C can be a non-volatile memory such as a ROM or flash memory. The boot memory 106C can store initialization code and, like the fixed storage 106A and language model cache 106B, the boot memory 106C can be communicatively linked to the processor modules 102

through the communications bridge 108. The boot memory 106C can be predominantly used during an initial power-on sequence at which time initialization code can be provided to the processor modules 102.

The communications bridge 108 can be an adapter to a host system such as a computer telephony (CT) system or VoIP gateway/endpoint. The communications bridge 108 can provide an interface to the local communications bus 104 through which data can be exchanged between the processor modules 102 and the host system.

Where the host system is a VoIP gateway/endpoint, an Ethernet switch (not shown) can be included which can process incoming and outgoing audio packets which conform to the VoIP protocol. In contrast, where the host system is a CT system, audio data can be received through a PCI interface.

In particular, where the local communications bus is a PCI bus and the host system provides a PCI interface, the communications bridge 108 can be a PCI-to-PCI bridge. Furthermore, where the host system is a CT system compliant with the Enterprise Computer Telephony Forum (ECTF) system architecture, the communications bridge 108 can be a PCI-to-PCI bridge having a PCI host interface 112 to the CT system and an audio interface 110 to an H.1x0 bus. In particular, where the speech processing board 100 includes a conventional PCI design, the audio interface 110 can be an interface to an H.100 bus. In contrast, where the speech processing board 100 includes a compact PCI (cPCI) design, the audio interface 110 can be an interface to an H.110 bus.

The communications bridge 108 also can include a processing element 114 for managing message communications between the speech processing board 100 and the host system according to a messaging protocol provided by the host system. 12. Finally, the speech processing board 100 can include an audio stream processor 116 coupled to the communications bridge 108. The audio stream processor 116 can manage incoming audio data arriving from either the audio interface 110 or the PCI interface 112. The audio stream processor 116 can be a programmable digital signal processor (DSP) and can be programmatically configured to extract audio data received in the communications bridge 108. Once extracted the audio information can be temporarily stored in local memory 118 before being distributed over a serial audio channels 122 to selected processor modules 102 by a local audio controller 120 based on hosted instances of speech application tasks in each processor module 102 as described by host system's messaging protocols.

II. Speech Processing Board Detail

The speech processing board 100 of the present invention can be logically viewed as having several subsystems including a communications subsystem (commands and data), a communications bridge, a processing subsystem, and a memory subsystem . In general, however, the speech processing board 100 has a basic method of operation which involves the execution of multiple instances of speech application task images such as speech recognition or TTS playback.

Communications Subsystem

In a preferred aspect of the present invention, the communications subsystem can include a PCI design that can be implemented in either standard PCI format or in cPCI format. The primary communications channel, PCI, is utilized by the communications bridge 108 to communicate specific commands and result sets stemming from those commands to and from the processor modules 102, to upload language models and finite state grammars to the storage system 106, and to upload firmware updates to the processor modules 102. Also, as will be apparent to one skilled in the art, audio data can be transferred to the speech processing board 100 both via the system PCI bus interface 112 and the audio interface 110.

By comparison, the local communications bus 104 can provide a communications path between processor modules 102. Additionally, the local communications bus 104 can serve as the communications medium between large vocabulary recognition tasks and corresponding language models. Notably, in one aspect of the present invention, the language models for use by speech recognition tasks in the speech processor board 100 generally can be stored in one of three system resources: the local memory of the processor modules 102, the local memory 118 of the communications bridge 108, or the fixed storage 106A.

Importantly, to minimize the response time of a speech recognition task, it can be helpful for the processor modules 102 to be able to access language models stored in the speech processor board 100 in as close to real-time as possible. For this reason, it is preferable that a local communications bus 104 is selected to be wider and faster than a corresponding host system bus. For example, one satisfactory configuration can

include a local communications bus 104 which is a 64 bit wide 133MHz PCI bus. This configuration yields a burst data rate which exceeds 1GB/s in throughput. Still, to facilitate the use of field programmable gate arrays (FPGAs) in the speech processor board 100, the local communications bus can be limited to a 64 bit wide 66MHz PCI bus yielding a maximum burst data rate of 528MB/sec.

Communications Bridge

In one aspect of the invention, the communications bridge 108 can be PCI-to-PCI bridge and can be included in a programmable logic block for instance an FPGA. The communications bridge 108 can include an audio interface 110 which can be configured to receive audio data from an H.1x0 bus. In particular, the audio interface 110 can be a bus end-point which is compliant with the ETCF hardware specifications H.100 (PCI) or H.110 (cPCI). Notably, the H.1x0 bus endpoint can be contained in a programmable logic block within the PCI-to-PCI bridge. Local memory 118 attached to the communications bridge 108 can serve as a communications buffer for audio data.

The programmable logic of the communications bridge 108 also can include a local audio controller 120 for local audio distribution. Specifically, audio data can be distributed over serial audio channels 122 which link the communications bridge 108 to the processor modules. The serial audio channels can be configured to communicate using conventional UARTs or I2C technology. Notably, recent revisions to the I2C interface can support 3.4 Mbps data streams.

Run-time commands and result sets can be passed through the host interface 112 between the speech processing board 100 and the host system. Typical runtime commands can include requests for the speech processing board 100 to perform an operation on a specific audio stream received through the audio interface 110 followed by command status responses, speech application task results and the like. For example, where a requested operation is a speech recognition task, the speech processing board 100 can report recognition results to the host system through the communications bridge 108. Optionally, the result sets can include associated probabilities.

In a preferred aspect of the invention, the communications bridge 108 includes the audio interface 110 through which audio streams can be communicated between the speech processing board 100 and the host system. Notwithstanding, where a host system does not support the H.1x0 bus, audio stream data can be provided through the host interface 112. In that case, the communications bridge 108 can detect the receipt of audio data and can route the audio data to an on-board audio communications function which can pre-process the audio data. Once pre-processed, the audio data can be routed to individual processing modules 102 as would be the case were the audio data received through the audio interface 110.

Processor Subsystem

Figure 2 is a block diagram of a processing module 102 for use in the speech processing board 100 of Figure 1. Specifically, the speech processing board 100 can

be configured either with commercial off-the shelf (COTS) processor modules, or processor modules specifically designed for performing speech processing tasks. In either case, each processor module 102 can include basic elements such as a CPU core 200 with on-board cache 202, local memory 204, local memory controller 206 and a processor local bus (PLB) 208 communicatively linking the core 200 with the controller 206. For instance, an exemplary processor module 102 can include a 555 MHz PowerPC core with 32K/32K instruction and data (I/D) caches, a 133MHz Processor Local Bus, 8KB of PLB-attached Static RAM (SRAM) and external SDRAM controllable by the core through a 64-bit PC-133/PC-266 Double Data Rate (DDR) SDRAM Controller.

The processor local bus 208 also can link the core 200 to an external communications bus such as the local communications bus 104 through a communications bridge 210 such as a PowerPC-to-PCI Interface Bridge. Notably, the processor module 102 can include on-chip ethernet channels. In consequence, the processor module 102 can be configured to directly transmit and receive audio packets from a VoIP gateway/endpoint over a packet-switched network.

As in the case of conventional processor modules, the processor module 102 of the present invention can include a DMA controller 212 such as a 4 Channel DMA Controller. Finally, the processor module 102 can include a serial interface 214, for example a v1.0 USB Controller and a 3.4 Mbps I2C interface, each accessible across the PLB 208 through a PLB to serial interface bridge 216. Notably, the entire processor

module 102 can be housed in a 404 I/O, 575 pin BGA package occupying approximately one square inch of board area on the speech processor board 100.

Memory Subsystem

The memory subsystem can be subdivided into a memory locally available in each processor module 102, and remote memory commonly available to each processor module 102. Locally, each processor module 102 can have various types of memory available for use by loaded speech application tasks including L1 I/D caches, on chip SRAM, and local high performance SRAM. Likewise, each processor module 102 can access remote SDRAM-based language model caches and remote bootstrap memory in non-volatile memory such as flash memory. The extended L1 cache sizes can be 32 KB. The on chip SRAM can be relatively small (less than 16 KB) and can be used primarily as a buffer for audio data exchanged with the local SDRAM. An on chip L2 cache can be optionally provided.

Local memory 204 can be addressed by an on chip local memory controller 206 that connects to the on chip processor local bus 208. In the case where the local memory controller 206 is a DDR SDRAM controller, the local memory controller 206 can support 266MHz DDR SDRAMs in 8 byte widths yielding burst data rates which exceed 2GB/sec. Notably, the data rates supported by a DDR SDRAM controller are substantially higher than conventional desktop computer memory designs and approximates the data rates of on chip L2 cache. In consequence, though an L2 cache can be included in a processor module 102, it is not required.

The local memory subsystem can provide a repository for speech application task program code, data tables, acoustic models, language model cache, complete finite state grammars, and memory structures associated with speech processing software. In addition, the local memory subsystem can include a portion allocated to a control program, for instance a real-time operating system (RTOS) which can manage memory allocation, task switching and communications activities. Significantly, a substantial portion of the local memory 204 of each processor module 102 can be allocated as a language model cache in order to further reduce traffic in the local communications bus 104. Similarly, finite state grammar tables can be stored locally in the local memory 204 of each processor module 102 having a loaded speech application task based thereon.

Three types of remote memory are available for use by processor modules 102 which can include boot memory 106C, a language model cache 106B, and the fixed storage 106A, each accessible via the communications bridge 108. The boot memory 106C can be accessed by the processor modules 102 during an initial power-on sequence. Specifically, once power has been applied to the speech processing board 100 or a bus reset has been detected, the communications bridge 108 can hold all of the processor modules 102 in a reset state. The reset can be deactivated to each processor module 102 which can issue a reset vector fetch directed to the boot memory 106C. The processor module then can load the RTOS and other initialization code into local memory 204, execute power-on diagnostics and enter an idle loop awaiting a command from the host system.

The language model cache 106B generally can include a complete image of one or more language models that are stored in the fixed storage 106A. Physically, the language model cache can be pluggable, volatile memory such as SDRAM configured in SO-DIMM packaging. In consequence, different memory configurations can be selected allowing for versions of the speech processing board 100 that are optimized for low cost, mainly small vocabulary tasks, or high performance NLU or large vocabulary tasks. Presently, the nominal SDRAM requirement for a single language large vocabulary can be 128MB while 32MB or less can suffice for systems utilizing sub-500 word, finite state grammar speech recognition tasks. Also, in the case where the local communications bus 104 is a 64 bit wide 133MHz PCI bus, the SDRAM can be 8 bytes wide and operate at 133MHz or 266MHz.

Importantly, the language model cache 106B can be mapped to a common address space where the language model cache 106 can be uniformly accessed by all processor modules 102 in the speech processing board 100. Specifically, as part of the initialization sequence performed by the speech processor board 100, individual language models can be loaded into volatile memory, for example SDRAM, according to a pre-defined memory schema. Each language model can be stored contiguously in memory. During the boot strap load process performed by each processor module 102, a uniform starting address can be provided to the processor module 102. Notably, in a preferred aspect of the present invention, only a small portion of the SDRAM is mapped into the host system memory address space as required for host communications.

The final memory type available for use by the processor modules 102 is the fixed storage 106A. The fixed storage 106A can be a compact device such as a Microdrive which can be linked to the communications bridge 108 via a CompactFlash (CF) controller similar to a PCMCIA IDE interface. One suitable CF controller for use with a fixed storage device such as the Microdrive has been manufactured by International Business Machines Corporation of Armonk, New York USA. The fixed storage 106A can store all active language models and finite state grammars in use by processor modules 102 in the speech processing board 100.

III. Integration of Speech Processing Board with ECTF Framework

The speech processing board 100 can provide speech processing services in one of several types of CT systems. To date CT systems have been generally proprietary implementations. Still, the Enterprise Computer Telephony Forum (ECTF) framework represents an effort to define a standard CT system architecture. The ECTF framework can reduce the complexity of integrating CT subsystems by defining general-purpose telephony components with fully specified interfaces to enable interoperability among different products from different vendors.

The ECTF framework references two types of servers. Application servers execute call control, administration, reporting, and media services applications in a distributed network. By comparison, CT servers provide the call control, administration, resource management functionality, network access, and media resources (lines, voice recognition, fax) required by the applications. Application servers and CT servers

communicate in client-server relationships. By thoroughly specifying the interfaces between application servers, CT servers, and the hardware and software components of each server, the broadest range of interoperability can be achieved.

The ECTF has developed a comprehensive CT Framework which encompasses:

5 Architecture, Modeling, Interfaces (Protocols and APIs) and ECTF Models. Often overlooked, models play an important role in a comprehensive framework of interoperability specifications. Models define the conceptual basis, terminology, and behaviors, and correct usage of interfaces. While interfaces define the syntax by which two components connect, models define the language.

10 The ECTF has defined the following models: C.001 Call Control Model, M.001 Administrative Services Model, S.100 Media Services Model, and R.100 Call Center Reporting Model. The ECTF also has defined the following interfaces: C.100 JTAPI Call Control, M.100 Administrative Services Interface, M.500 SNMP MIB Specification, S.100 Media and Switching Services Interface, S.200 Transport Protocol Interface, 15 S.300 Service Provider Interface, S.410 JTAPI Media Interface, H.100 CT Bus for PCI, and the H.110 CT Bus for Compact PCI.

Figure 3 illustrates a CT architecture based on an ECTF framework which incorporates the speech processing board 100 of the present invention. Specifically, Figure 3 is a schematic illustration of the speech processing board 100 of Figure 1 20 integrated with a generalized ECTF-compliant CT media services system 300. The media services system 300 can process CT media services applications to share media resources and integrate with existing call control architectures. Media services refers to

the branch of CT technology that is concerned with media processing, including playing and recording of voice files, speech recognition and text-to-speech technology, DTMF detection and generation, and T.30 and T.611 fax services. Media services technology involves making media processing resources in a telephone system available to client software.

The media services system 300 can include a CT hardware layer 302, resource modules 304, a service provider interface 306 to system services modules 308, protocol interface 310, and an application programming interface 312 to CT applications 314. The media services system also can include a call control module 316 and a call control API 318 providing access to the call control module 316 for call control applications 320. Notably, the speech processing board 100 can integrate with the media services system 300 at the service provider interface 306.

In general the media services system 300 assumes that the speech functions are independent engines which receive audio streams and respond with speech recognized text. The routing of the audio stream, specification of related grammars and vocabularies are the responsibility of a call routing stack. This set of functions includes identifying the level of speech application support required to support the call which can be pre-defined based on a number called and the state of the call.

All grammars, vocabularies, acoustic and language models are assumed to be resident on the speech processing board 100 and pre-loaded into the language model cache 106B based on the defined set of speech application tasks. Additional copies of the various data sets for other inactive tasks, including different languages, generally

can be resident on the fixed storage 106A. Task management tools accompanying the speech processing board 100 can assist users in defining grammars and conversational models. These tools can tag the appropriate data sets resident on the fixed storage 106A for loading into the language model cache 106BA as needed.

IV. Conclusion

The ECTF model provides a straightforward entry point for the speech processing board 100 in a CT environment since all of the call management software can be used generally except that some modifications may be necessary to recognize that multiple levels of speech application functionality can be supported. In this manner the speech processing board 100 can focus on execution of instances of speech application tasks, on board audio path management on a per task basis, and management of host messaging protocols.

The present invention can be realized in hardware, software, or a combination of hardware and software. Moreover, the present invention can be realized in a centralized fashion in one computer system, or in a distributed fashion where different elements are spread across several interconnected computer systems. Any kind of computer system - or other apparatus adapted for carrying out the methods described herein - is suited. A typical combination of hardware and software could be a general purpose computer system with a computer program that, when being loaded and executed, controls the computer system such that it carries out the methods described herein. The present invention can also be embedded in a computer program product,

which comprises all the features enabling the implementation of the methods described

herein, and which when loaded in a computer system is able to carry out these

methods. Computer program means or computer program in the present context

means any expression, in any language, code or notation, of a set of instructions

5 intended to cause a system having an information processing capability to perform a particular function either directly or after either or both of the following a) conversion to another language, code or notation; b) reproduction in a different material form.

Significantly, this invention can be embodied in other specific forms without departing from the spirit or essential attributes thereof, and accordingly, reference should be had to the following claims, rather than to the foregoing specification, as indicating the scope of the invention.